



HAL
open science

Making data more available and usable for the scientific community: the case of the French Center for Socio-Political Data

Alina Danciu, Alexandre Mairot

► To cite this version:

Alina Danciu, Alexandre Mairot. Making data more available and usable for the scientific community: the case of the French Center for Socio-Political Data. 9th International Conference on Social Science Methodology, The International Sociological Association's (ISA) research committee RC33 on Logic and Methodology, Sep 2016, Leicester, United Kingdom. hal-02874810

HAL Id: hal-02874810

<https://sciencespo.hal.science/hal-02874810>

Submitted on 3 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Making data more available and usable for the scientific community

The case of the French Center for Socio-Political Data (CDSP)

RC33 2016
9th International Conference on Social Science Methodology
Wednesday, 14th of September, Leicester, United Kingdom

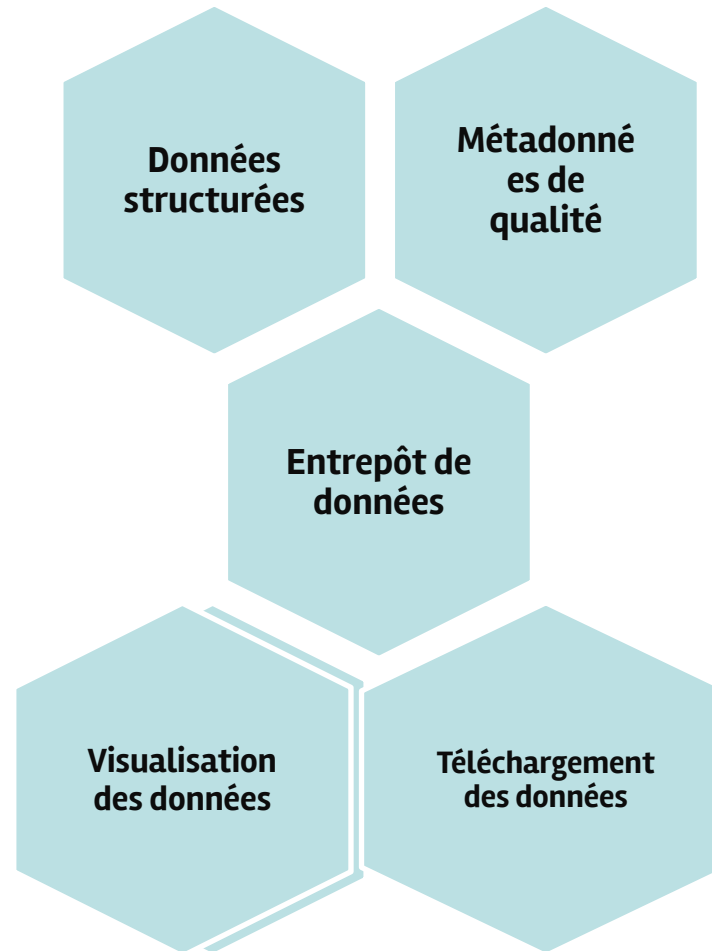
Alina Danciu, CDSP
Alexandre Mairot, CDSP

Secondary research

- More and more **data intensive** and **collaborative**
- The **amount of data** collected, analysed and stored has increased **enormously**
- **Finding** the data you need can be rather **difficult**



Des données accessibles



The “power” of DDI

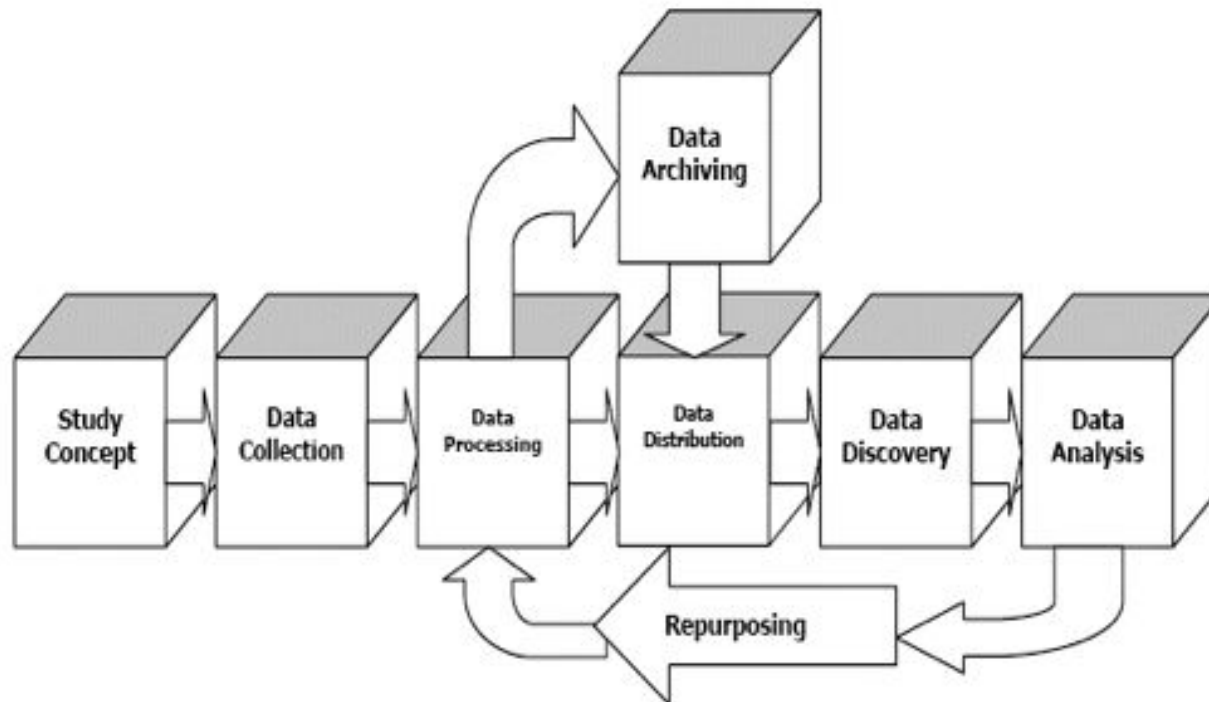


Figure 3: Combined Life Cycle Model

Le Centre de données socio-politiques (CDSP)

- Le CDSP diffuse des données quantitatives (2005) et qualitatives (2012)



Thèmes

- Comportements politiques /46
- Culture /3
- Éducation /1
- Elections /142
- Elus /9
- Environnement /5
- Famille /4
- Genre /1
- Groupes sociaux /1
- Institutions politiques /5
- Opinions politiques /47
- Professions /6
- Santé /11
- Science /5
- Sexualité /9
- Territoire /26
- Travail /2

Les relations entre générations au prisme des normes de solidarité et de justice sociale
Enquête quantitative - 2014

Fécondité, contraception, dysfonctions sexuelles (2013)
Enquête quantitative - 2013

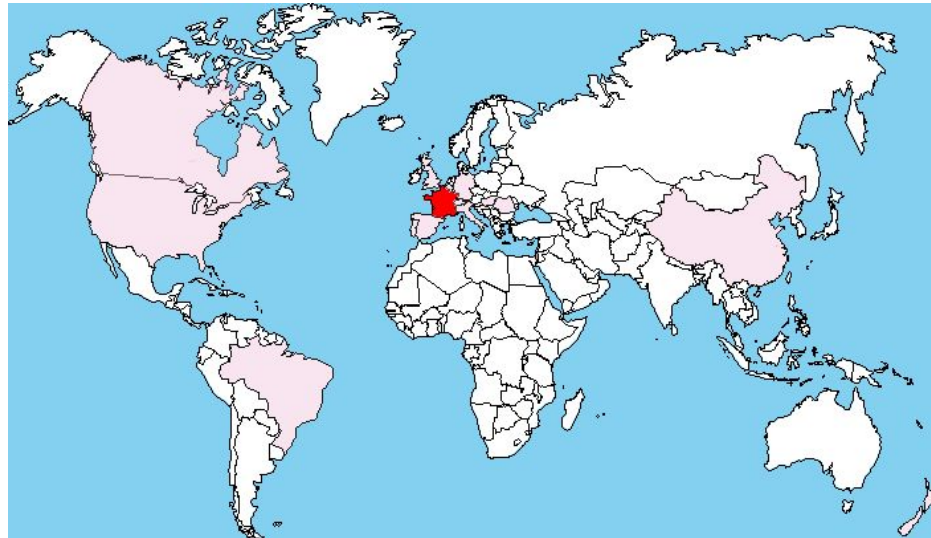
Pratiques culturelles, médias et technologies de l'information - vague 1 (2013)
Enquête quantitative - 2013

Dynamiques de mobilisation - vague 1 (2013)
Enquête quantitative - 2013

Enquête annuelle ELIPSS - vague 1 (2013)

Qui sont les utilisateurs de nos données ?

Geographical provenance:



Source: CDSP Quetelet portal users, 2015

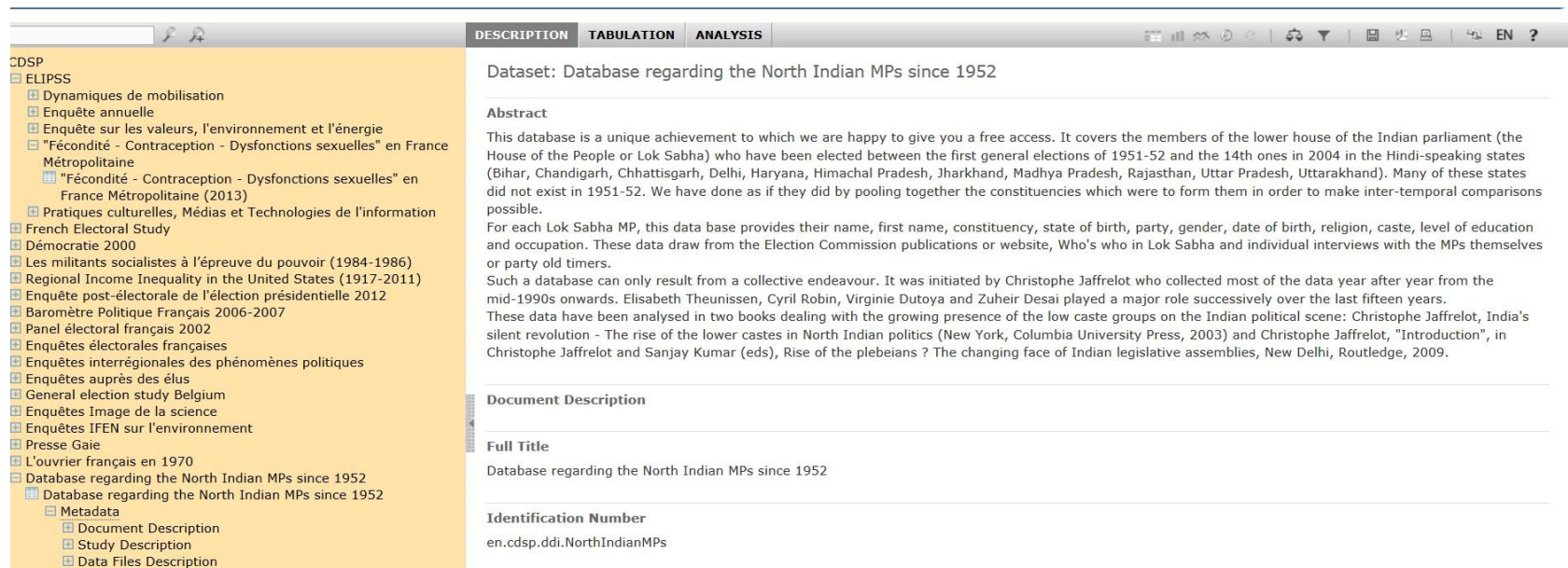
Disciplines:

Sciences politiques (45%), économie (25%), sociologie(25%)

Statuts:

chercheurs (50%), doctorants (25%), étudiants (25%)

Visualisation de données quantitatives : Nesstar



The screenshot shows the Nesstar interface with a sidebar on the left containing a tree view of datasets under the CDSP category. The main content area is divided into three tabs: DESCRIPTION, TABULATION, and ANALYSIS. The 'DESCRIPTION' tab is active, displaying the following information for the dataset 'Database regarding the North Indian MPs since 1952':

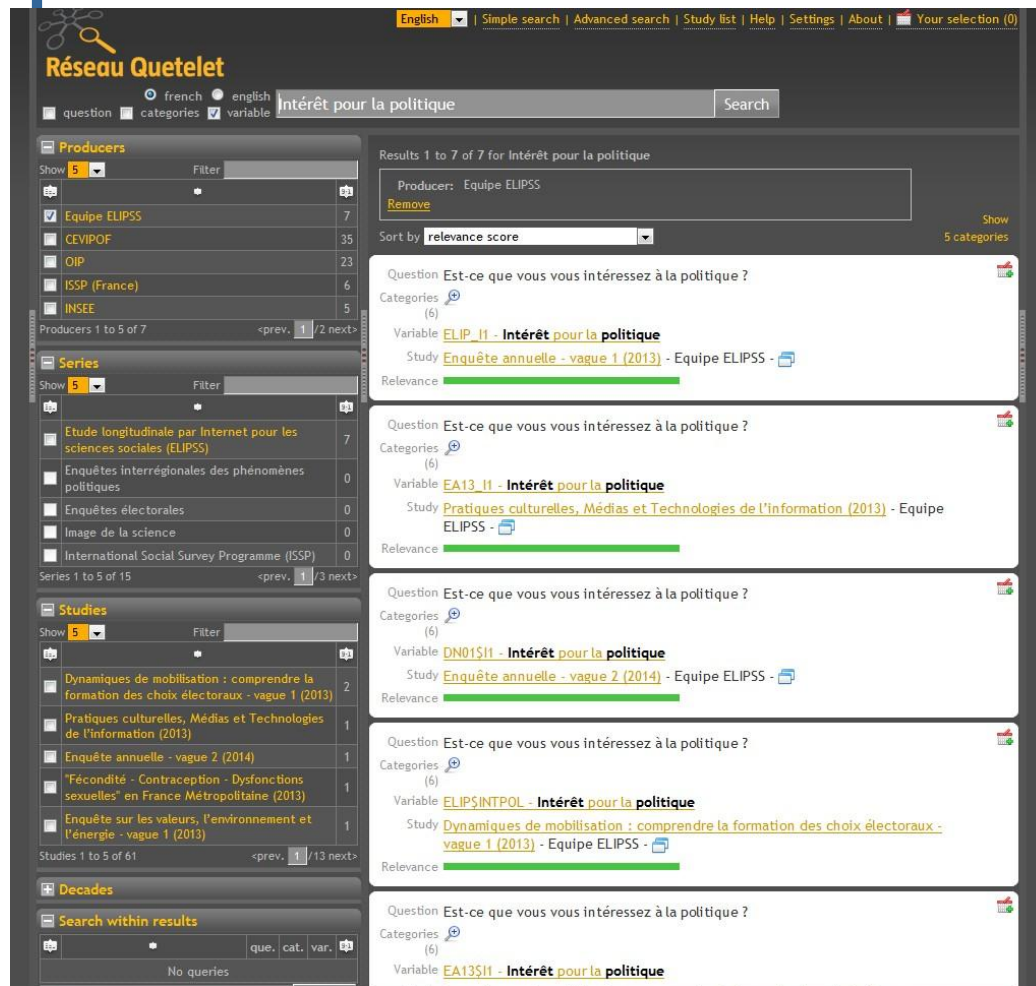
- Dataset:** Database regarding the North Indian MPs since 1952
- Abstract:**

This database is a unique achievement to which we are happy to give you a free access. It covers the members of the lower house of the Indian parliament (the House of the People or Lok Sabha) who have been elected between the first general elections of 1951-52 and the 14th ones in 2004 in the Hindi-speaking states (Bihar, Chandigarh, Chhattisgarh, Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh, Uttarakhand). Many of these states did not exist in 1951-52. We have done as if they did by pooling together the constituencies which were to form them in order to make inter-temporal comparisons possible.

For each Lok Sabha MP, this data base provides their name, first name, constituency, state of birth, party, gender, date of birth, religion, caste, level of education and occupation. These data draw from the Election Commission publications or website, Who's who in Lok Sabha and individual interviews with the MPs themselves or party old timers.

Such a database can only result from a collective endeavour. It was initiated by Christophe Jaffrelot who collected most of the data year after year from the mid-1990s onwards. Elisabeth Theunissen, Cyril Robin, Virginie Dutoya and Zuheir Desai played a major role successively over the last fifteen years. These data have been analysed in two books dealing with the growing presence of the low caste groups on the Indian political scene: Christophe Jaffrelot, India's silent revolution - The rise of the lower castes in North Indian politics (New York, Columbia University Press, 2003) and Christophe Jaffrelot, "Introduction", in Christophe Jaffrelot and Sanjay Kumar (eds), Rise of the plebeians ? The changing face of Indian legislative assemblies, New Delhi, Routledge, 2009.
- Document Description:**
- Full Title:** Database regarding the North Indian MPs since 1952
- Identification Number:** en.cdsp.ddi.NorthIndianMPs

Visualisation de données quantitatives : Base de questions



The screenshot displays the Réseau Quetelet database interface. At the top, there is a search bar with the query "Intérêt pour la politique" and a "Search" button. The interface is divided into several sections:

- Producers:** A list of data producers with their respective counts. "Equipe ELIPSS" has 7 items, CEVIPOF has 35, OIP has 23, ISSP (France) has 6, and INSEE has 5.
- Series:** A list of data series. "Etude longitudinale par Internet pour les sciences sociales (ELIPSS)" has 7 items, while others like "Enquêtes Interrégionales des phénomènes politiques" and "Enquêtes électorales" have 0 items.
- Studies:** A list of studies. "Dynamiques de mobilisation : comprendre la formation des choix électoraux - vague 1 (2013)" has 2 items, while others have 1 item.
- Search within results:** A section for refining the search with no queries currently applied.
- Results:** The main area showing 7 results for the query. Each result includes the question text, categories, variable name, study name, and a relevance score bar.

The search results are sorted by "relevance score". The first result is for the question "Est-ce que vous vous intéressez à la politique ?" with variable "ELIP_11 - Intérêt pour la politique" and study "Enquête annuelle - vague 1 (2013) - Equipe ELIPSS".

Téléchargement de données : portail Quetelet

[My account](#) [My requests](#) [New request](#) [My messages](#)

Studies selection

▼ Your research project [Edit my research project](#)

Test

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis sit amet pulvinar nisi, convallis fringilla nisi. Sed at pellentesque ante, sit amet bibendum risus. Sed varius consectetur lacus eu laoreet. Fusce euismod ligula nec nisl tempor vulputate. Mauris sodales feugiat egestas. Vestibulum maximus ornare hendrerit. Ut congue metus a lacus volutpat, non iaculis neque efficitur. Suspendisse potenti. Suspendisse id sapien consectetur, molestie magna in, aliquam magna.

Nunc fringilla hendrerit venenatis. Vestibulum hendrerit purus quis dolor lobortis ornare. Morbi ultrices justo vitae elit interdum, in sagittis sapien ullamcorper. Nam feugiat, neque non eleifend pharetra, risus mi interdum massa, ut egestas ante odio in turpis. Cras laoreet, lacus at ornare ultricies, velit odio sollicitudin nunc, et pulvinar odio mauris ullamcorper mi. Curabitur purus nunc, maximus eu venenatis eu, convallis non elit. Aenean ut lorem purus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam fermentum felis elit. Nam scelerisque vehicula metus, sit amet facilisis urna porttitor bibendum. Aliquam non suscipit diam, sed semper nibh.

No attachment found

All distributors ADISP CDSP INED
Selected studies 1

Search :
(Delete)
 Reference
 Title
 Producer
 Year
 Search

Information : page 1 of 18, 10 studies of 171

Title	Producer	Year	Diffusers	Reference	
■ Baromètre CIDEM de la citoyenneté, vague 1	CIDEM (Civisme et démocratie)	2001	CDSP	cdsp_cidem_v1	» more info
■ Baromètre CIDEM de la citoyenneté, vague 2	CIDEM (Civisme et démocratie)	2002	CDSP	cdsp_cidem_v2	» more info
■ Baromètre CIDEM de la citoyenneté, vague 3	CIDEM (Civisme et démocratie)	2002	CDSP	cdsp_cidem_v3	» more info
■ Baromètre CIDEM de la citoyenneté, vague 4	CIDEM (Civisme et démocratie)	2002	CDSP	cdsp_cidem_v4	» more info
■ Baromètre CIDEM de la citoyenneté, vague 5	CIDEM (Civisme et démocratie)	2003	CDSP	cdsp_cidem_v5	» more info

Lost?



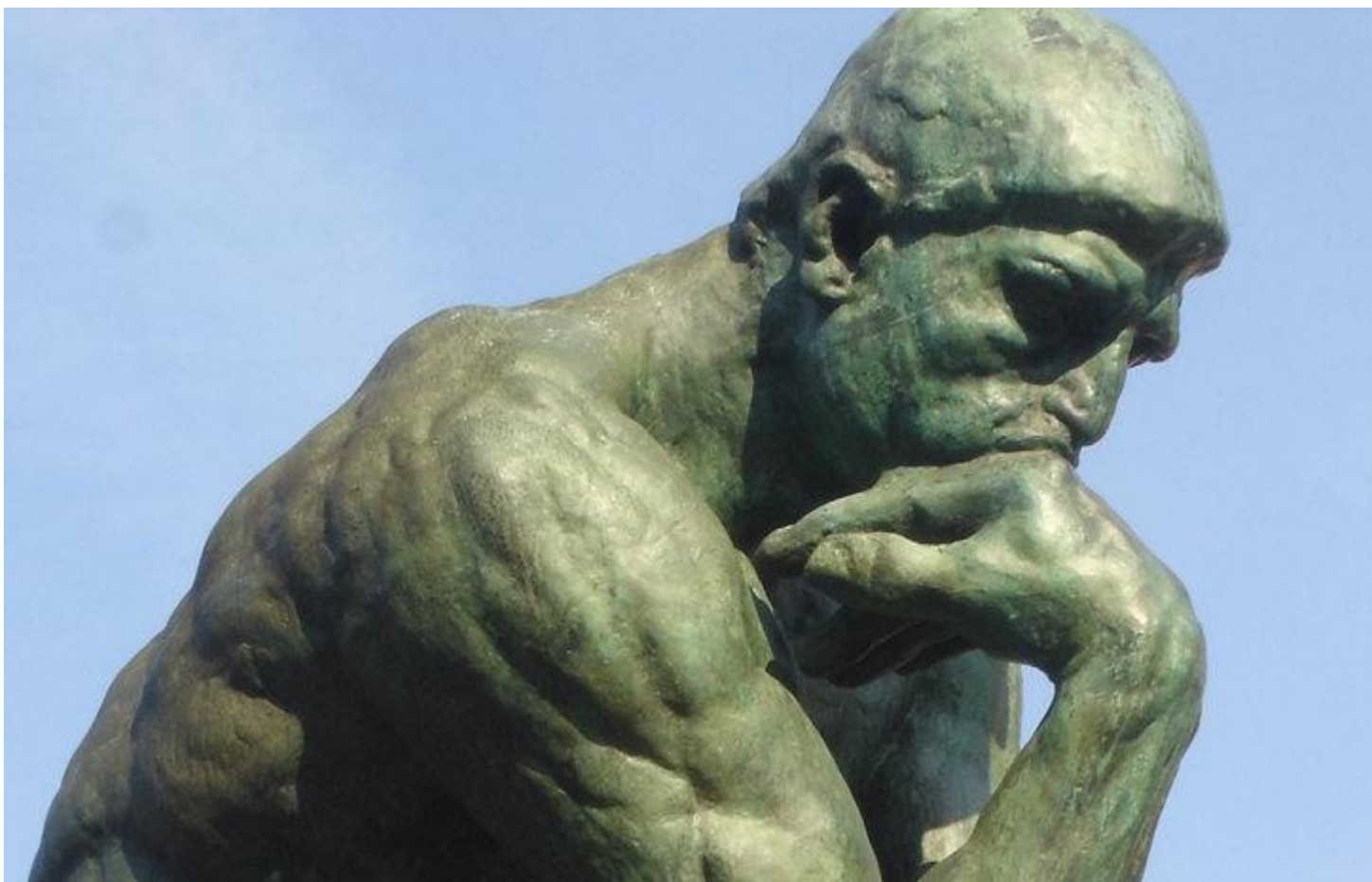
Finding the appropriate data

A crucial step of a research project based on secondary analysis

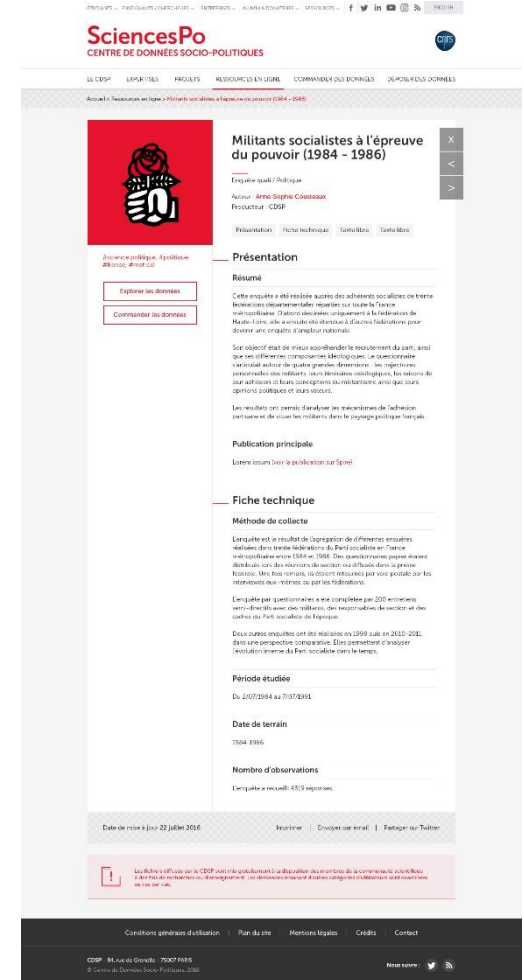
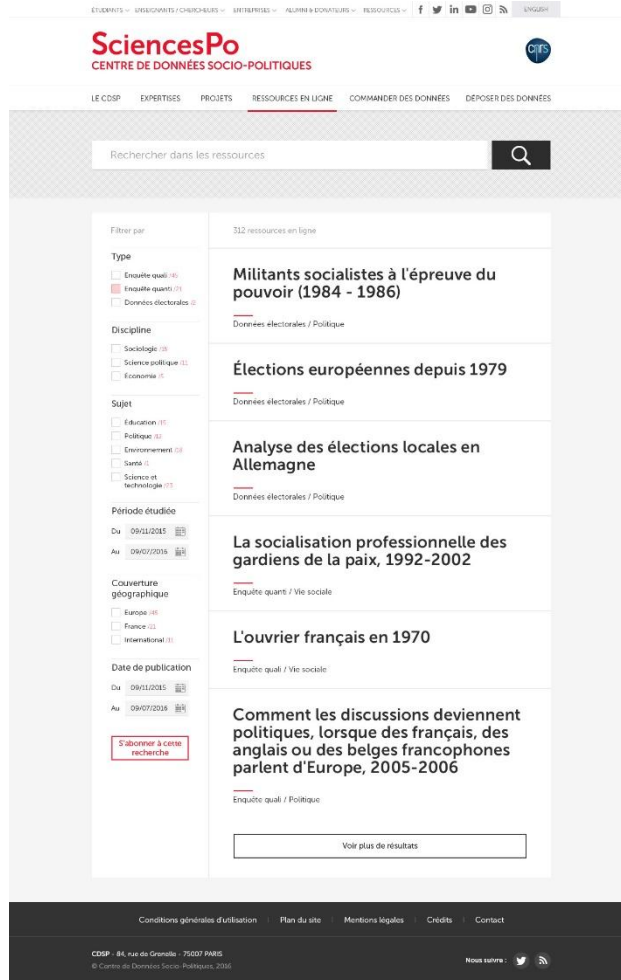
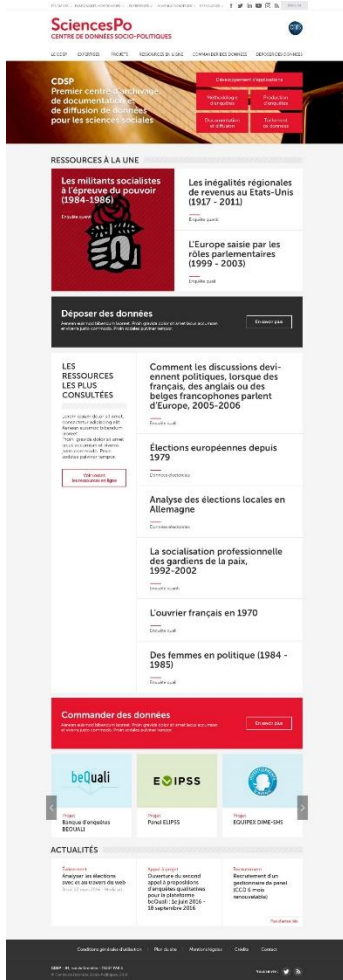
but :

- Studies and variables difficult to find
- Metadata not entirely harmonised across waves/studies
- Several **CDSP dissemination platforms** not *yet* inter-operable

How to improve data discovery and data download?



A new CDSP website



Variable harmonisation

- **Litteral question:** How interested would you say you are in politics? Are you?
- **Variable categories**
 - ✓ 1- Very interested
 - ✓ 2- Quite interested
 - ✓ 3- Hardly interested
 - ✓ 4- Not at all interested

Study	Variable
Image de la science	Int (2001)
Baromètre Politique Français 2006-2007	Q3
Panel électoral français 2002	Xq201
Le baromètre de suivi de l'environnement IFEN-OIP	Q12 (1993), q26 (2000)
Enquête interrégionale des phénomènes politiques	Q29 (1998), q26 (1999), q17 (1994), q11 (1992)
Enquête annuelle ELIPSS	ELIP\$11 (2013), DN01\$11(2014)
Pratiques culturelles, Médias et Technologies de l'information (ELIPSS)	EA13\$11

- **Variable label:** How interested in politics

Data more available and usable

- **Approach to favour data discovery:** data harmonisation across waves and studies
- **Visualisation** of all the waves including the question and download the ones of interest
- **Datamatching upon request**

A new approach including DDI-Lifecycle

- 2014
 - ✓ How to integrate the DDI-Lifecycle specifications in our dissemination practices?
 - Using DDI-Lifecycle tools?
 - Dealing with DDI-Codebook documented data?
- 2015
 - ✓ How about the creation of our metadata system?
 - An exploration of social sciences data centers' portals
 - An analysis of the current situation
 - Decision to build a metadata system
- 2016
 - ✓ Process of building a **metadata database**

Procedures

- ✓ Metadata de-duplication
- ✓ Versioning metadata
- ✓ Integration of concepts

- ✓ A new discovery data portal inspired by the DDI Daghestul workshop 2014 and 2015
- ✓ Definition of the web tools and technology
- ✓ ...

Thank you for your attention !

alina.danciu@sciencespo.fr
alexandre.mairot@sciencespo.fr